

Modeling Unobserved Heterogeneity in Contingent Valuation of Health Risks

Arana, Jorge E.; León, Carmelo

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:

Arana, J. E., & León, C. (2006). Modeling Unobserved Heterogeneity in Contingent Valuation of Health Risks. *Applied Economics*, 38(19), 2315-2325. <https://doi.org/10.1080/00036840500427460>

Nutzungsbedingungen:

Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: <http://www.peerproject.eu> Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under the "PEER Licence Agreement". For more Information regarding the PEER-project see: <http://www.peerproject.eu> This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft



Modeling Unobserved Heterogeneity in Contingent Valuation of Health Risks

Journal:	<i>Applied Economics</i>
Manuscript ID:	APE-05-0012
Journal Selection:	Applied Economics
Date Submitted by the Author:	11-Jan-2005
JEL Code:	Q51 - Valuation of Environmental Effects < Q5 - Environmental Economics < Q - Agricultural and Natural Resource Economics, C25 - Discrete Regression and Qualitative Choice Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, D62 - Externalities < D6 - Welfare Economics < D - Microeconomics, I19 - Other < I1 - Health < I - Health, Education, and Welfare
Keywords:	Contingent Valuation, Flexible Distributions, Heterogeneity, Health Risk, Willingness to Pay

MODELING UNOBSERVED HETEROGENEITY IN CONTINGENT VALUATION OF HEALTH RISKS

Abstract

Human preferences for alternative levels of health risks can be heterogeneous. In this paper we consider a flexible distribution approach to model health values elicited with the dichotomous choice contingent valuation method. Rigid parametric structures cannot model sample heterogeneity while imposing strong assumptions on the error distribution. We consider a mixture of normal distributions which can approximate arbitrary well any empirical distribution as the number of mixtures increases. The model is applied to data on willingness to pay for reducing the individual risk of an episode of respiratory illness. The mixture distribution model is compared with the rigid probit model using a Bayes factor test. The results show that the mixture modeling approach improves performance while allowing for the consideration of alternative groups of individuals with different preferences for health risks.

Keywords. Contingent Valuation, Flexible distributions, Heterogeneity, Health preferences, Normal mixture, Willingness to pay.

1
2
3 **1. Introduction**
4

5
6 The economic valuation of health risks is relevant for adopting efficient health policies, and
7
8 is increasingly approached with non-market valuation methods such as contingent valuation
9
10 (CV). This is a survey approach involving a market transaction carried out on a
11
12 questionnaire posed to the relevant population. Past studies by Viscusi et al. (1991) and
13
14 Johannesson et al. (1996) among others have shown that people are sensitive to the level of
15
16 risk of illness.
17

18
19
20
21 However, the values people place on policies for reducing health risks could be subject to
22
23 unobserved heterogeneity. This can be explained because of a number of related reasons: i)
24
25 subjects can have different degrees of risk aversion, ii) they could perceive and/or be
26
27 sensitive to health risks in different ways, and iii) respondents could present different
28
29 preferences for alternative health statuses. These sources of variability in the data can rarely
30
31 be observed by the researcher.
32
33

34
35
36 Thus, sample data on the values of health risks elicited following the CV method can be
37
38 conditioned by unobserved heterogeneity, breaking down the assumption of the same
39
40 distribution across the sample. This assumption is maintained in parametric and semi-
41
42 parametric models for the dichotomous choice elicitation method. This is a method which
43
44 requests binary responses from the subject to a given price for the proposed policy
45
46 involving the reduction in health risks. Most applications in health valuation have fitted a
47
48 parametric distribution such as the probit model to these binary responses.
49
50

51
52
53 In this paper we consider a mixture distribution approach which allows us to represent
54
55 heterogeneity across the sample in contingent valuation surveys utilizing the dichotomous
56
57
58
59
60

choice format. The proposed approach has both theoretical and practical advantages. From the statistical point of view, it provides flexibility and with enough components virtually any empirical distribution can be approximated arbitrarily well (Zubov (1985)). In particular, multiples modes are possible and fatter tails than the normal can also be accommodated by mixing in normal components with large variance. From the economic point of view, unlike most of the other methods, an estimator of the link function may be strictly monotonic. The violation of this assumption implies that consumers are not necessarily utility maximizing (Coppejans, 1999).

There are other approaches which have been proposed for flexible modeling of this type of data. For instance, Horowitz (1996) proposes a consistent and asymptotically normal nonparametric estimator for single index models with unknown transformation of the dependent variables. However, this estimator it is not necessarily monotonic and these results do not extend to the binary choice model. Another possible way to incorporate flexibility in the model is by using a exponential generalized beta type 2 (Nagler, 1994). Of particular interest, is the work of McDonald (1996, 2000), who extended this family of beta distributions to permit heterogeneity in the scale parameter. McDonalds' results suggest that a random coefficient probit performs well compared with semiparametric alternatives.

In the context of modeling discrete choice contingent valuation (DCCV) data (see Hanemann and Kanninen (1999) for a detailed review). This is the case of the flexible semi-nonparametric model by Creel and Loomis (1997), which is similar to Chen and Randall (1997). These approaches allow the researcher to deal with the misspecification of rigid distribution functions imposed on irregular empirical data, but do not address the problem of unobserved heterogeneity across the sample observations. In addition, by using

Monte Carlo simulations, Araña and León (forthcoming) show that the mixture of normals approach reduces bias and improves performance with respect to the Creel and Loomis semi-nonparametric model, particularly when the sample is characterized by heterogeneous preferences. Another nice feature about the mixture of normals is that it can be viewed as a natural extension of the paradigm case for purely parametric setting – the probit model.

The mixture model is applied to data obtained from the valuation of the reduction in the probability of an episode of respiratory illness. These data are likely to show heterogeneity because of the different degrees of risk aversion and different risk perceptions that people can have with respect to potential health outcomes. The model for sample heterogeneity based on a mixture of normal distributions is shown to improve performance with respect to a rigid probit model, which is also rejected according to Horowitz and Hardle’s (1994) test.

The outline of the paper is as follows. Section 2 presents the modeling approach to unobserved heterogeneity and the estimation algorithm using Bayesian Markov chain Monte Carlo. Section 3 discusses an extension of the Horowitz and Hardle’s (1994) test for probit specification. Section 4 presents the data obtained from a CV survey on the reduction of the risk of an episode of respiratory illness, while section 5 discusses the results. Finally, Section 6 summarizes the main conclusions.

2. Model of sample heterogeneity

The dichotomous choice format requests from the subject to answer a yes/no willingness to pay question. The subject is offered a bid price (B_j) for a reduction in the level of health risk

from z^0 to z^1 . Let $e_i(\cdot)$ be the expenditure function for individual i , that is, the inverse of the indirect utility function with respect to income. If the subject seeks the largest level of utility then the answers would be yes if $e_i(z^1, V^*) + B_j \leq e_i(z^0, V^*)$ and no otherwise, where V^* is some fixed level of utility, and we assume other arguments of the expenditure function, such as prices, as constants.

The expenditure difference could be seen as the individual's willingness to pay (WTP) for the reduction in health risk, that is,

$$WTP_i = e_i(z^1, V^*) - e_i(z^0, V^*). \quad (1)$$

Therefore, the observed answer $\{y_i\}$ to bid price $\{B_i\}$ takes the value one if WTP_i is higher or lower than the bid price, and zero otherwise. We can assume that the latent variable WTP is a function of two components, a deterministic μ and a random component ξ . In general, we can write $WTP_i = WTP_i(\mu_i, \xi_i)$, where μ is the mean of WTP and ξ_i is a random error term, which is assumed normally distributed, with zero mean and σ standard error. Assuming independent answers and fixed covariates, the probability of a positive answer to bid price B_i is

$$\Pr ob(y_i = 1) = \Pr ob[WTP_i(\mu_i, \xi_i) > B_i] = F_{\xi} [WTP_{i|\xi}^{-1}(B_i, \mu_i)] \quad (2)$$

where $\mu_i = x_i' \beta$ is the linear predictor associated with a $k \times 1$ regression parameter vector β and a covariate vector x_i , including z , and $WTP_{i|\xi}^{-1}$ is the inverse of the willingness to pay function with respect to ξ_i . The linear predictor is linked to the probability of a positive response by a known cumulative distribution function $\{F_{\xi}\}$ or link function.

The error distribution can be specified as some parametric form, and the model estimated by maximum likelihood (Hanemann and Kanninen, 1999)). However, if there are different groups of individuals across the sample this assumption cannot be maintained, since each group could have a specific distribution, and unimodality does not hold. Thus, the assumption of a common distribution can be overcome by considering that the stochastic terms adopts the following form:

$$\xi_i = e_{i1}(\alpha_1 + \eta_1 \sigma_1) + e_{i2}(\alpha_2 + \eta_2 \sigma_2) + \dots + e_{ik}(\alpha_k + \eta_k \sigma_k) \quad (3)$$

where α_j and σ_j ($\forall j=1,2,\dots,k$) represent the mean and the standard deviation of each of the normal forms in the mixture¹, and $\alpha' = (\alpha_1, \dots, \alpha_k) \in R^k$; $\sigma' = (\sigma_1, \dots, \sigma_k) \in R_+^k$ and $\eta_i | X \sim N(0, 1)$. The random vectors $e' = (e_1, \dots, e_k)$ are i.i.d. with multinomial probability distribution of parameters $\pi_j = \text{Prob}(e_{ij} = 1)$ ($j=1, \dots, k$).

Without further restrictions, the model is clearly unidentified in the sense that more than one set of values of the parameters imply the same $p(y|X)$. In particular, to prevent interchanging the components of the mixture some labelling restrictions are needed. Here we impose i) $\alpha_1 < \alpha_2 < \dots < \alpha_m$. The additional restrictions require are the followings²: ii) $\text{rank}(X)=k$ and $a'X' \neq (1, \dots, 1)$ for any $k \times 1$ vector a ; (ii) $p_j > 0 \forall j$; (iii) the support of $x_i' \beta$ is a set of positive Lebesgue measure; (v) $\sigma_j = 1$ for some j .

¹The mean and the variance of the error terms are $E[\xi_i] = \sum_{j=1}^k \pi_j \alpha_j$; $Var[\xi_i] = \sum_{j=1}^k \pi_j^2 \sigma_j^2$ respectively,

and where $\sum_{j=1}^k \pi_j = 1$

² See Geweke and Keane (1999) for specific details about the specification.

The inclusion of a mixture of normals is also a way of introducing flexibility in the distribution assumption, since it deviates from the assumption of a single parametric distribution for the entire sample. Thus, the mixture model approaches a semi-parametric model, since it eliminates the normality assumption. In addition, Heckman and Singer (1984) and Zubov (1995) show that any distribution can be approximated arbitrarily well by increasing the number of mixtures k in the model.

The likelihood function across the sample under the mixture assumption is:

$$p(y|\beta, \alpha, \sigma, p, X) = \prod_{i=1}^N y_i \sum_{j=1}^k \pi_j \Phi[(B_i - \alpha_j - x_i' \beta) / \sigma_j] + \\ + (1 - y_i) \left\{ 1 - \sum_{j=1}^k \pi_j \Phi[(B_i - \alpha_j - x_i' \beta) / \sigma_j] \right\} \quad (4)$$

Maximum likelihood estimators (MLE) of the parameters can be obtained by accordingly optimizing this function using some non-linear method. However, sampling distributions of MLE cannot be derived in closed form so that estimator properties rely almost exclusively on asymptotic or large sample situations. In particular, any model requires extremely large (as much as 1000 observations per parameter) samples to insure the adequacy of asymptotic approximates (McCulloch and Rossi (1994)). In addition, asymptotic properties of ML estimators do not need to be maintained with small and finite samples³. Anderson and Richardson (1979) and Griffiths, Hill and Pope (1987) found out relevant biases with numerical simulations of probit and logit models with small samples, while Copas (1988)

³ See Amemiya (1985) or Huber (1981) for a more detailed discussion of the asymptotic properties of ML estimators.

utilize Taylor series expansions to define the bias obtained for a logit model with small samples.

In these cases it could be utilized at Bayesian approach (see for example Araña and León (2002)). Bayesian methods, as developed by Chib (1992) and Albert and Chib (1993), adhere to the likelihood principle and are conducted using formal rules of probability theory. This means that under mild conditions Bayes estimators are consistent, asymptotically efficient and admissible in small samples. As a practical matter, Bayesian inference is free from the use of asymptotic approximations and delivers exact, finite sample inference (Rossi and Allenby (2002)). The prior distributions are defined as follows:

$$\beta \sim N(\underline{\beta}, \underline{H}_{\beta}^{-1})$$

$$\alpha \sim N(\underline{\alpha}, \underline{H}_{\alpha}^{-1})$$

$$\underline{s}_j^2 / \sigma_j \sim \chi^2(\underline{v}_j)$$

$$p(1/\sigma) = \prod_{j=1}^k \left[\frac{1}{2^{\underline{v}_j/2} \Gamma(\underline{v}_j/2)} \right] \left(\underline{s}_j^2 \right)^{\underline{v}_j/2} \frac{1}{\sigma_j^{(\underline{v}_j-2)/2}} \exp(-0.5 \underline{s}_j^2 / \sigma_j)$$

$$p(\pi) \sim \text{Beta}(r)$$

where $r = (r_1, \dots, r_k)$, $\alpha = (\alpha_1, \dots, \alpha_k)$ and \underline{H}_{α} is an $k \times k$ matrix. The last two matrixes are positive defined. The posterior distribution is derived by using a Gibbs sampling algorithm, similar to Geweke and Keane (1999). This involves sampling from the conditional posterior distributions when they are known in a feasible form. Thus, even though $\text{WTP} = (\text{WTP}_1, \text{WTP}_2, \dots, \text{WTP}_n)'$ is not observed, it is possible its simulation from available information.

Thus, given $Y = (y_1, y_2, \dots, y_n)'$ and $\theta = (\beta, \alpha, \sigma, \pi)$, the posterior distribution following data augmentation $\pi(\theta|Y, WTP)$ and the conditional density of the latent variable $f(WTP_i|Y, \theta)$ are known in a manageable form⁴.

Taking the starting value for θ i.e. $\theta^{(0)} = (\beta^{(0)}, \alpha^{(0)}, \sigma^{(0)}, \pi^{(0)})$, the Gibbs sampling algorithm obtains iterated samples from each of the posterior conditional distributions. The algorithm is carried on t times leading to the simulated vector $(WTP^{(t)}, \beta^{(t)}, \alpha^{(t)}, \sigma^{(t)}, \pi^{(t)})$ obtained from the joint distribution $(WTP, \beta, \alpha, \sigma, \pi)|Y$. These series of algorithms of size t are repeated over H times, leading to H values for each parameter which are simulated from the posterior distribution, i.e. $[WTP_h^{(t)}, \beta_h^{(t)}, \alpha_h^{(t)}, \sigma_h^{(t)}, \pi_h^{(t)}]_{h=1}^H$. The moments of interest are obtained from these simulated values.

3. A formal test of the probit model

Before applying the flexible modeling approach of a mixture distribution it is convenient to test whether a rigid model such as probit is appropriate for the binary choice data obtained in a CV experiment. Horowitz and Hardle (1994) propose a specification test (H-H) for the distributional assumption of the error term of the probit model. In this section we adapt this test to the parameterization of a latent variable WTP model.

The test is based on the difference between the parametric fit of the model $\Phi(X'\beta)$ and the non-parametric regression of Y over $X'\hat{\beta}$. If the link function is correctly specified then

⁴ The explicit conditional posterior distributions are presented in appendix 2.

this difference should be due only to sampling errors. The statistics under the null hypothesis is as follows:

$$T_n = \sqrt{h} \sum_{i=1}^n u(X_i' \hat{\beta}) [Y_i - \Phi(X_i' \hat{\beta})] [\tilde{F}_i(X_i' \hat{\beta}) - \Phi(X_i' \hat{\beta})] \quad (5)$$

which is asymptotically normal with zero mean and variance σ_T^2 ⁵. The value of h determines the bandwidth of the confidence interval in the kernel regression. $u(\cdot)$ is a weighted function which gives smaller relevance to extreme observations, and $\tilde{F}_i(\cdot)$ is a kernel regression on the data with order $r \geq 2$. Parameter β is estimated under the null assumption using ML for the probit model. $\tilde{F}_i(\cdot)$ should have the properties of being asymptotically unbiased and uncorrelated with Y_i . The first property is obtained by using Bierrens (A.6) correction. The second follows by eliminating observation i from the computation of the kernel regression or “leave-one-out” estimation. The asymptotic variance σ_T^2 is replaced by the consistent estimator. We utilize a quartic kernel, which satisfies the requirements for Horowitz⁶ proposition 1 to hold:

$$K(v) = \frac{15}{16} (1 - v^2)^2 I(|v| \leq 1) \quad (6)$$

where I is the indicator function. This estimator is asymptotically unbiased. Thus there is no trade-off between asymptotic bias and variance as is commonly the case with standard techniques of determining the bandwidth. Proença (1993) used Monte Carlo simulation to show that the test can be influenced by the choice of h with finite samples. These results

⁵ Appendix 1 presents the development of the statistics in more detail.

⁶ See for instance, Martins (2001) for a demonstration.

suggest that under-smoothing could leads to the rejection of H_0 . Thus, it is convenient to utilize large values of the bandwidth in order to improve the potency of the test.

4. Application

The data for the application of the mixture distribution model come from a contingent valuation study conducted in the island of Gran Canaria (Canary Islands) to measure the economic benefits from the reduction in the probability of an episode of respiratory illness caused by air pollution. The field work was carried out in 1999. The questionnaire was improved after two pre-tests studies and two focus groups where the key design aspects were discussed. The final sample was obtained from 368 in-person interviews addressed to the general population of the island. The sample was taken randomly in the households subject to quotas for representative age and sex groups. Protest responses as indicated by individuals who opposed the valuation process because the government should pay for the risk prevention program amounted to 10.5 percent of the sample. Those subjects who were not willing to pay any positive quantity represented another 11.2 percent.

The valuation scenario presented the subject with a hypothetical reduction in the possibility of becoming ill with an episode of respiratory illness. The wording of the scenario is shown in Appendix 3. The symptoms of this type of health problem were described by a card which specified also the number of days of their duration. The subjects were informed that the episodes could be associated with some diseases such as asthma and bronquitis. The possible causes of the episodes were attributed to traffic pollution in some areas of the city of Las Palmas. The possibilities of having an episode could also increase due to whether

conditions, particularly in the days of the year when Alisios winds stop blowing through the Canary Islands. The possibility of developing the episode of respiratory illness could be reduced by implementing policy measures involving traffic restrictions in the days of higher risks.

Heterogeneity in the sample could emerge because some people might not be interested in reducing the possibilities of a respiratory illness. This behavior can be explained because they could not feel as potentially affected by this health problem. That is, they could perceive that their risk of getting the health problem is low because of their current good health conditions or because they have already taken precautionary measures. In addition, the proposed solution might involved unpopular measures such as the restrictions in the use of private transportation that some people could prefer against the reduction in the probability of illness in the general population.

As a provision rule, the subject was asked to assume that the proposed measures were going to be undertaken only if the majority of the population agreed on them. The subject was also recalled about her budget constraint and was asked about her willingness to pay for the reduction in the probability of having an episode of respiratory illness. The suggested risk reduction levels were presented with the aid of a risk ladder and were defined as 20, 40, 60 and 70 percent reductions. The risk card was tested in focus groups and was found effective as a risk communication device.

The elicitation format was single bounded dichotomous choice involving a single binary yes/no question to a given price for a concrete risk reduction level. Each subject received up to four binary choice questions for the alternative risk reduction levels which were

randomly ordered with the aim of avoiding order bias effect. The price vector was designed with the distribution of the responses to an open ended question in the pretest study and following Cooper's (1993) design method for a predetermined number of bids. The bid price vector was 6, 12, 18 and 24 Euros for the lowest risk level, which was proportionately raised for the higher levels of risk reduction.

5. Results

The H-H test allows us to determine whether there is need to utilize a more flexible modeling approach to the empirical data. This could be convenient in order to avoid misspecification errors as coming from the utilization of rigid parametric structures, such as probit or logit. The use of these models could result in biased estimates of the structural parameters defining the mean and median willingness to pay for the reduction of the risks of respiratory illness.

The H-H test statistics was computed for different values of h , in order to compare the sensitivity of the results to the choice of this parameter. The variance was estimated by (A.5) and the weighting function $u(v)$ was assumed to take the unitary value inside the confidence interval and zero in other case (Proença, 1993). The limits of the intervals were taken as the 10th and 90th percentiles, although the results were not sensitive to changes in these limits.

Table 1 presents the results of the test statistics for different values of h and their p -values. The probit model is not rejected for low values of h . However, for values of h above 0.55 the test leads to the rejection of this rigid specification at the 5 percent level, suggesting that

there is need to consider a more flexible approach to model the empirical data. The left tail of the estimated distribution is always fatter for any value of h than the non-parametric approach. Thus, the probit model particularly fails to represent those subjects on the tails of the distribution. A more flexible approach might be able to model more precisely these extreme responses to the WTP question, establishing different classes of individual positions with respect to the value of the health prevention policy.

Since we have no prior information on model parameters for the Bayesian models, we assume very non-informative diffuse priors with a large variance. The starting values for Gibbs sampling were taken from ML estimation, although the results were quite robust to changes in these parameters. The posterior results were generated by running the Markov chains for a burn-in period of 10000 draws and the retaining every 10th draw of the next 150000 draws. Convergence checks as in Raftery and Lewis (1992) and Geweke (1992) did not indicate any problems, and this was corroborated by visual inspection of trace plots and rerunning the samplers many times from different starting values.

The estimated WTP model for the reduction of the risk of respiratory illness includes some covariates which allowed us to improve the predictive power and the goodness of fit of the model. Table 2 presents the description of the covariates which were significant in explaining WTP, obtained after probing with various potential specifications. The results of the flexible Bayesian model for heterogeneous samples are presented in Table 3.

For comparison purposes we also present the results of the Bayesian probit model (BPM) estimation following Albert and Chib's (1993) approach applied by Araña and León (2002) in DCCV, which disregards modeling heterogeneity. There are three main reasons for this

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

decision. Firstly, simulation studies show that under non informative priors the Bayesian probit model (BPM) produces asymptotically equivalent results that the random coefficients probit model (RCPM) (Geweke and Keane, 1999). The random coefficients models have been shown to be an interesting alternative to semiparametric and nonparametric approaches to incorporate heterogeneity (McDonald, 1996). Secondly, the asymptotic properties of the RCPM estimated by maximum likelihood are not guaranteed with small and finite samples, while the BPM was shown to improve the results with respect to maximum likelihood estimation, especially for small samples. Finally, since the BPM is the most simply version of the mixture of normals model, the comparison of its fit with the more complex mixture models can be used as an additional test of the relevance of the unobserved heterogeneity in a specific application (Bianchi, 1997).

WTP is significantly influenced by income, as expected by theory. The effect of income is quadratic, indicating that WTP raises with income but eventually at a decreasing rate. On the other hand, WTP is also explained by the number of days of the potential episodes of respiratory illness and by the level of risk reduction. Both variables have positive effects on WTP. That is, those subjects posed with larger number of days of the episodes and larger risk reduction levels responded with higher latent WTP values. In addition, WTP becomes smaller for those subjects that have some type of chronic disease. This result could seem surprising, but can be explained because the marginal benefit of a health risk reduction measure could be smaller for the already ill.

The value of k for the number of normal mixture distributions was chosen by the best Akaike and Bayesian information criteria statistics. The result was $k=4$, i.e. the best model was obtained with a mixture of four distributions reflecting four latent segments in the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

population regarding the preferences for the health risks prevention measures. The four latent classes differ particularly in mean WTP. There are three classes with positive means while there is one class with a large negative mean WTP. The class with the largest mean value (latent segment 3) is willing to pay 41.70 Euros for the risk prevention measures. The other two classes with positive values (latent segments 1 and 2) are quite close in their mean WTP, although there is less uncertainty in segment 2 due to its shorter confidence interval.

Negative WTP is reflected in latent segment 4 with a mean WTP of -31.78 Euros. These are people who might object the adoption of preventive measures for the risk of respiratory illness, probably because they oppose the traffic restrictions or because they think they are not affected by the contaminated air. Overall, mean WTP with the BMNP model is practically equal to the one obtained with the basic model that does not take into account unobserved sample heterogeneity. However, it is clear that the policy implications are quite different. Whereas in the basic model every segment of the population would be treated the same, in the flexible model the population could be grouped into different classes, each charged with its corresponding mean WTP.

The flexible model also gives us the sample proportions for each latent segment. It can be seen in Table 3 that the largest proportion is obtained for the segment with negative WTP (38% of the sample) while the smallest proportion is found for those subjects who are willing to pay the smallest positive amount (12%). The subjects who are willing to pay the largest amount to prevent respiratory illnesses represent 29% of the sample.

The results with the BMNP model improve substantially over the basic model. It can be seen that the t-ratio statistics become larger when the latent classes are modeled. In order to compare both models we calculate the Bayes' factor. This is defined by the ratio of the predictive probabilities of the alternative models, and can be interpreted as the times that one model is more likely than an alternative model.

The likelihood of alternative models can be evaluated with the predictive distribution or marginal likelihood. In general, for a model k with data density distribution $\pi_k(y|\theta_k)$ and prior density $\pi_k(\theta_k)$, $\theta_k \in \Theta_k$, the predictive probability is $\pi_k(y) = \int \pi_k(y|\theta_k) \pi_k(\theta_k) d\theta_k$, which is the denominator of Bayes' theorem. If we want to compare two alternative models, for instance a probit model with $k=1$ versus a four mixture model with $k=4$, then the Bayes factor is defined as the ratio of their respective predicted probabilities, i.e. is π_{k4}/π_{k1} . In the case of equal prior probabilities for each model, the Bayes' factor becomes $B_{k4,k1} = \pi_{k4}(y|\theta_{k14}) / \pi_{k1}(y|\theta_{k1})$. For any particular value θ^* of the parameters, the predictive likelihood can be obtained from Bayes' theorem as $\ln(\pi(y)) = \ln(\pi(\theta^*)) + \ln(\pi(y|\theta^*)) - \ln(\pi(\theta^*|y))$. Hence, the predictive likelihood can be evaluated by combining the values of the prior, the likelihood and the posterior density at any value of θ^* .

The results of the computed marginal likelihoods or predictive distributions for both the probit model and the mixture model are presented in Table 3. The larger probability for the mixture model means that the Bayes' factor favors the mixture model against the more restricted probit model. Thus, the mixture model adds to the explanation of the latent WTP data, although it increases the parameterization of the model. The implication is that modeling unobserved heterogeneous preferences results in better predictions and more

accurate estimation of health risk benefits. That is, health preferences for the risk of a respiratory illness are subject to sample heterogeneity. The assumption of a common distribution structure is not maintained with these data and would lead to biased results of the welfare estimates of the health policy measures.

6. Conclusions

Health preferences could be heterogeneous because of the different perceptions people in society might have about their health status, their environment and their probability of becoming ill. The level of risk aversion to particular health states could also vary across individuals, leading to heterogeneity in preferences. The result is that the values elicited with health valuation methods such as contingent valuation can not be assumed to be attributed to a unique representative individual. Rather, there might be different patterns of behavior across the sample which are reflected in different health values. Rigid parametric approaches to model health valuation empirical data impose strong assumptions, since all individuals are treated as emerging from the same data generation process.

In this paper we have considered a mixture of normal distributions to model unobserved heterogeneity with the dichotomous choice contingent valuation method. The model is flexible because any empirical distribution can be represented by increasing the number of mixture distributions. Flexible approaches can accommodate different patterns of behavior resulting from the dispersion of individuals' preferences. In order to simplify computations, the mixture model is estimated utilizing Bayesian methods involving data augmentation and Gibbs sampling. The results of the mixture model are compared with the probit model

with the Bayes' factor, showing that it leads to a larger predictive probability under the assumption of a four mixture distribution.

The model has been applied to data on the valuation of the reduction of the probability of a respiratory illness potentially caused by air pollution or natural factors. The mixture model selects four different distributions embedded within the data, which might respond to different preferences for health risks. The group with the largest proportion is given by those individuals who are not willing to pay any amount or might be on the negative side when asked for an amount of money for the reduction of the risk of respiratory illness. These people might have developed negative preferences based on their perceptions of the potential risks associated with environmental pollution. Thus, the mixture distribution approach is particularly appropriate to model zero and negative values in situations where there might be potential conflicts of interest or different opinions about the implementations of the health prevention measures.

Subjects who were willingness to pay positive amounts also showed different degrees of intensity in their preferences, with the largest proportion represented by the group with the higher positive values. The mixture model improves prediction against the model which imposes a unique distribution on the data. The latter structure is also rejected by utilizing the H-H test for probit misspecification, calling for a more flexible approach to model the data. Thus, the mixture approach reduces bias in the estimation of the benefits from reducing health while providing a more accurate representation of the different preferences within society. Further research should focus on the development of practical methods to identify the characteristics of these groups of individuals. In addition, the mixture model

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

could be extended to other stated preference methods commonly applied in the valuation of health preferences.

For Peer Review

References

Albert, J.H. and Chib, S. (1993). "Bayesian Analysis of Binary and Polichotomous Response Data". *Journal of American Statistical Association*, 88, pp. 669-679.

Amemiya, T. (1985). "Advances Econometrics". Harvard University Press. Cambridge.

Anderson, E.B. and Richardson (1979). "Logistic Discrimination and Bias Correction in Maximun Likelihood Estimation". *Technometrics*, 21, 71-78.

Araña, J.E. and Leon, C.J. (2002). "Willingness to pay for Health Risk Reduction in the context of Altruism". *Health Economics* vol 11. Iss.7. pp-623-635.

Bianchi, M. (1997). "Testing for Convergence: Evidence from Non-Parametric Multimodality Tests". *Journal of Applied Econometrics*, 12(4), pp. 393-409.

Bierens, H. (1987). "Kernel Estimators of Regression Functions". *Advances in Econometrics: Fifth World Congress*, Vol 1, Bewley RF (eds). Cambridge Uniersity Press. New York. 99-144.

Chen, H.Z. and Randall, A. (1997). "Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation". *Journal of Econometrics*, 76, pp-323-40.

Chib, S. (1992). "Bayes Inference in the Tobit Censored Regression Model". *Journal of Econometrics*, 51, pp. 79-99.

Cooper, J.C. (1993). "Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys". *Journal of Environmental Economics and Management*, 24, 25-40.

Copas, J.B. (1988). “Binary Regresión Models for Contaminated Data” . Journal of the Royal Statistical Society, B, 50:2, 225-265.

Coppejans, M. (2001). “Estimation of the Binary Response Model using a Mixture of Distribution Estimator”. Journal of Econometrics, 102, 231 –269.

Creel, M. And Loomis, J. (1997). “Semi-Nonparametric Distribution-Free Dichotomous Choice Contingent Valuation”. Journal of Environmental Economics and Management, 32, 341-358.

Geweke, J. and Keane, M. (1999). “Mixture of Normals Probit Models” en Hsiao, C., Lahiri, K., Lee, L.F. and Pesaran, M.H. (eds). *Analysis of Panels and Limited Dependent Variables: A Volume in Honor of G.S. Maddala*. Cambridge University Press. Cambridge.

Geweke, J.F. (1992). “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments” en *Bayesian Statistics*. Edited by J. Bernardo, J. Berger, A. Dawid F. Smith. Oxford University Press, New York, 169-193.

Griffiths, W.H., Hill, R.C. and Pope, P.J. (1987). “Small Sample Properties of Probit Models Estimators”. *Journal of the American Statistical Association*, 82, pp 929-937.

Hanemann, M.W. and Kanninen B. (1999). “The Statistical Analysis of Discrete-Response CV Data”. Working Paper N° 798.

Heckman, J. and Singer, B. (1984). “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data”. *Econometrica*, 52(2), 271-320

Horowitz, J.L. (1996). “Semiparametric estimation of a regression model with an unknown transformation of the dependent variable”. *Econometrica* 64,103 –137.

Horowitz, J.L. and Härdle, W. (1994). "Testing a Parametric Model Against a Semiparametric Alternative". *Econometric Theory*, 10, 821-848.

Huber, P.J. (1981). "Robust Statistics". New York. John Wiley and Sons.

Johannesson, M., Johansson, P-O., O'Connor, R.M., (1996). "The Value of Private Safety Versus the Value of Public Safety". *Journal of Risk & Uncertainty*, 13, pp 263-275.

Martins, M.F.O.(2001). "Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal". *Journal Of Applied Econometrics*, vol. 16, iss. 1, pp 23-39.

McDonald, J.B. (1996). "An application and comparison of some flexible parametric and semi-parametric qualitative response models". *Economic Letters*, 53, pp. 145-152.

McCulloch, R.E. and Rossi, P.E. (1994). "An Exact Likelihood Analysis of the Multinomial Probit Model". *Journal of Econometrics*, 64, pp. 0217-228.

Nagler, J. (1994) "Scobit: an alternative estimator to logit and probit". *American Journal of Political Science*, 38, pp. 230- 255.

Proenca, I. (1993). "On the Performace of the HH Test ". Working paper 9319, Institu de Statistique, UCL.

Raftery, A.E. and Lewis, S.M. (1992). How many iterations in the Gibbs sampler? in *Bayesian Statistics 4* (J.M. Bernardo et al., editors), Oxford University Press, pp. 763-773

Rossi, P.E. and Allenby, G.M. (2002). "Bayesian Statistics and Marketing". Working Paper Ohio State University. Department of Marketing & Logistics.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Titterington, D.M., Smith, A.F.M and Makov, U.E. (1985). "Statistical Analysis of Finite Mixture Distributions". Wiley, New York.

Viscusi, W.K., Magat, A. and Huber, J. (1991). “Pricing Environmental Health Risks: Survey Assessments of Risks-Risks and Risks-Dollar Trade-Offs for Chronic Bronchitis”. *Journal of Environmental Economics and Management*, 21, pp 32-51.

Zubov,V.I.. (1995). “Standardization of the analytic representation of probability distributions and of stochastic matrices”. *Russian Academy of Sciences, Doklady, Mathematics*, 50,164 –168.

Appendix 1. H-H test statistic

Horowitz and Hardle (1994) propose a procedure to test for rigid structures for the error distribution in endogenous dichotomous choice models. In general, the problem is to test the specification of the following model:

$$E[Y|X = x] = f(x, \beta) \quad (\text{A.1})$$

where $Y \in \mathbb{R}$, $x \in \mathbb{R}^k$ and f is a function in \mathbb{R} which is known and associated to the parameter vector $\beta \in \Theta \subset \mathbb{R}^k$. Let us consider a random sample $\{(x_1, y_1), \dots, (x_n, y_n)\}$ where (x_i, y_i) , $i = 1, \dots, n$, is jointly distributed $g(x, y|\beta)$ defined in $\mathbb{R}^k \times \mathbb{R}$. The general regression function $f(x, \beta)$ can be considered a simple index model, such as probit, where $E[Y|X = x] = P(Y = 1|X = x) = \Phi(x'\beta)$. When equation (A.1) is correct, we have that $E[Y - f(x, \beta)|x] = 0$. Thus, the functions of the explanatory variables should be correlated with the errors $Y - f(x, \beta)$. If we assume that function $f(x, \beta)$ can be written as a function F of $v(x, \beta)$, where F and v are known functions defined in \mathbb{R} , then equation (A.1) takes the following form:

$$E[Y|X = x] = F[v(x, \beta)] \quad (\text{A.2})$$

If we have the simple index model $v(x, \beta) = x'\beta$ and function $F[v(x, \beta)] = F(x'\beta)$, where F is the link function, then the test involves the comparison of the following hypotheses:

$$H_0 : E[Y|X'\beta = v] = F(v)$$

$$H_1 : E[Y|X'\beta = v] = H(v)$$

where H is an unknown function. When the link function is probit, the test statistic is as follows:

$$T_n = \sqrt{h} \sum_{i=1}^n u(X_i' \hat{\beta}) [Y_i - \Phi(X_i' \hat{\beta})] [\tilde{F}_i(X_i' \hat{\beta}) - \Phi(X_i' \hat{\beta})] \quad (\text{A.3})$$

Under the null hypothesis and some regularity conditions, it can be shown that T_n is asymptotically distributed $N(0, \sigma_T^2)$ with

$$\hat{\sigma}_T^2 = 2C_K \int_{-\infty}^{+\infty} u^2(v) [\sigma^2(v)]^2 dv \quad (\text{A.4})$$

where $C_K = \int_{-\infty}^{+\infty} k(u)^2 du$ and $\sigma^2(v) = \text{Var}(Y|X'\beta = v)$. A consistent estimator for σ_T^2 in endogenous dichotomous choice variables is given by

$$\hat{\sigma}_T^2 = \left(2C_K / n \right) \sum_{i=1}^n \left\{ u(X_i' \hat{\beta})^2 [\tilde{F}_h(X_i' \hat{\beta}) [1 - \tilde{F}_h(X_i' \hat{\beta})]]^2 \right\} / \hat{p}_h(X_i' \hat{\beta}) \quad (\text{A.5})$$

The weighted function $u(v)$ can be set to one if it is inside a bounded interval and zero in other case.

In order to obtain Bierrens's (1987) property of consistency, $\tilde{F}(v)$ can be defined as a linear combination of Kernel regressors with different bandwidth (h and s respectively). This correction relies on the following result:

$$\sqrt{nh} \{ \tilde{F}(v) - F(v) \} \sim N(0, C_K \sigma^2(v) / p_h(v)) \quad (\text{A.6})$$

where $p_h(v)$ is the density function for v estimated by $p_h(v) = (nh)^{-1} \sum_i \{(v - v_i)/h\}$.

Appendix 2. Gibbs Sampler

The conditional posterior densities over which Gibbs sampling is applied are the following:

$$f(WTP_i | Y, \theta) = \begin{cases} \phi(WTP_i | x_i' \beta + \alpha' e_i, \sigma_{L_i}^{-1}) I[B_i, \infty] & \text{si } y_i = 1 \\ \phi(WTP_i | x_i' \beta + \alpha' e_i, \sigma_{L_i}^{-1}) I[-\infty, B_i] & \text{si } y_i = 0 \end{cases} \quad (\text{A2.1})$$

$$\pi(\beta | Y, WTP, \alpha, \sigma, \pi) = \phi\left(\beta \middle| \bar{\beta}, \bar{H}_\beta^{-1}\right) \quad (\text{A2.2})$$

$$\pi(\alpha | Y, WTP, \beta, \sigma, \pi) = \phi\left(\alpha \middle| \bar{\alpha}, \bar{H}_\alpha^{-1}\right) \quad \text{s.t. } \alpha_1 < \alpha_2 < \dots < \alpha_k^7 \quad (\text{A2.3})$$

$$f(\bar{s}_j^2 / \sigma_j | Y, WTP, \beta, \alpha, \pi) = f_{\chi^2}(\bar{s}_j^2 / \sigma_j | \bar{\nu}_j) \quad \text{s.t. } \sigma_1 < \sigma_2 < \dots < \sigma_k^6 \quad (\text{A2.4})$$

$$f(\pi | Y, WTP, \beta, \alpha, \sigma) = f_{Beta}(\pi | r_1 + T_1, r_2 + T_2, \dots, r_k + T_k) \quad (\text{A2.5})$$

where $\phi(\cdot) I[a, b]$ is the normal pdf truncated in the interval $[a, b]$, $T_j = \sum_{i=1}^N e_{ij}$ and f_{χ^2} , f_{Beta}

represents chi-square and beta pdf's respectively. Then, the vector of means are

$$\bar{\beta} = \bar{H}_\beta^{-1} \left(\bar{H}_\beta \beta + \sum_{i=1}^N x_i WTP_i / \sigma_{L_i} \right); \text{ and } \bar{\alpha} = \bar{H}_\alpha^{-1} \left[\bar{H}_\alpha \alpha + \sum_{i=1}^N e_i (WTP_i - x_i' \beta) \right].$$

Thus, $L_i = (i: e_{ij}=1)$; $P(L_i = j) = \pi_j \exp\left[-(WTP_i - \alpha_j - x_i' \beta)^2 / 2\sigma_j\right]$ and the covariance

matrix are $\bar{H}_\beta = \left(\underline{H}_\beta + \sum_{i=1}^N h_{L_i} x_i x_i' \right)$ and $\bar{H}_\alpha = \left(\underline{H}_\alpha + \sum_{i=1}^N e_i e_i' \right)$. Then,

$$\bar{s}_j^2 = \underline{s}_j^2 + \sum_i e_{ij} (WTP_i - \alpha_j - x_i' \beta)^2; \text{ and } \bar{\nu}_j = \underline{\nu}_j + T_j.$$

⁷ This restriction guarantees the identifiability of the model (Geweke and Keane, 1999)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix 3

In the next question we want to know how much people value a reduction in the possibility that the population of Las Palmas have of becoming ill because of a respiratory problem. The following card shows the symptoms of an episode of respiratory illness, which could be associated with some diseases such as asthma or bronchitis.

These types of episodes are possible in the city of Las Palmas because of traffic pollution, and tend to increase in some times of the year when there are adverse climatic conditions, such as when the North Atlantic Alisios winds stop blowing through the Canary Islands. A possible solution which could reduce the incidence of the episodes of respiratory illness would be to implement traffic restrictions in the days that traffic emissions could cause more problems. The reduction in the possibility of having an episode of respiratory illness depends on the type and effectiveness of the measures which are undertaken. Let us considered that this possibility could be reduced in 20%, 50% or 70% as shown in the following card.

Assume that measures involving traffic restrictions in some areas are undertaken only if they benefit to the majority of the population. Now consider how much you value in money terms the reduction in the possibility of becoming ill of a respiratory illness.

Also recall all the personal damages that these symptoms would imply to you, including the costs on medicines and working time lost, as well as your limited income and your necessary expenses. Would you be willing to pay x euros for the reduction in the possibility of become ill in x %?

Yes No Do not know/ Refuse answer

Table 1. Horowitz and Härdle (H-H) test results

	$h = 0.3$	$h = 0.5$	$h = 0.55$	$h = 0.6$	$h = 0.7$	$H = 1$
Statistics	1.587	-0.606	-1.198	-1.780	-2.904	-5.740
p-value	0.9432	0.2721	0.1159	0.0374	0.0028	0.0001

Table 2. Description of explanatory variables in WTP models.

RP	= 1 for a 20% risk reduction, =2 for a 40% risk reduction, =3 for a 60% risk reduction and =3.5 for a 70% risk reduction
HS	Self-reported health status index for the individual (1-4), where 1 is very good and 4 is very bad.
AGE	
INCOME	Annual personal income of the subject in Euros.
EXERCISE	= 1 if individual usually does exercise at least twice a week, 0 otherwise
CHRONIC	= 1 if individual has a chronic disease, 0 otherwise.

Table 3. Descriptive Statistics

<i>Variable</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Min</i>	<i>Max</i>
RP	2.3315	0.8213	1	4
CHRONIC	0.6277	0.4840	0	1
DAYS	5.6311	2.8543	4	7
INCOME	1116.07	436.14	0.4808	2.4040
HS	2.9103	0.7158	1	4
AGE	36.5815	10.6616	14	72
EXERCISE	0.56521	0.4964	0	1

Table 4. WTP valuation functions (standard error in parentheses)

<i>Variables</i>		Bayesian normal (k=1)	BMNP(k=4)
DAYS		31.90 (12.85)	23.87 (4.87)
RP		20.0472 (9.53)	18.67 (7.10)
CHRONIC		20.35 (12.08)	18.93 (10.53)
INCOME		0.0355 (0.0243)	0.0203 (0.01)
HS		-5.77 (8.81)	-6.06 (6.99)
AGE		-2.80 (1.37)	-2.11 (0.55)
EXERCISE		11.69 (11.58)	10.45 (9.83)
σ		96.10 (46.06)	80.71 (11.99)
Latent segment 1	Mean WTP	-	36.54 [19.97, 57.60]
	Proportion	-	0.38
Latent segment 2	Mean WTP	-	0
	Proportion	-	0.09
Latent segment 3	Mean WTP	-	-12.45 [-49.86, 24.96]
	Proportion	-	0.12
Latent segment 4	Mean WTP	-	96.28 [59.11, 133.44]
	Proportion	-	0.41
Total population	Mean WTP	71.79 [28.06, 154.85]	51.86 [24.23, 79.55]
N		368	368

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 5. Model Selection based on Marginal Likelihood values (the preferred model is highlighted in bold)

	<i>Marginal Likelihood</i>
K=1	-2366.83
K=2	-1898.14
K=3	-1836.56
K=4	-1822.68
K=5	-1832.37